A box-jenkins sarima analysis of dysentery cases in children aged below five at a provincial hospital in Zimbabwe

Dr. Smartson P. Nyoni[1], Mr. Thabani Nyoni[2]

[1]ZICHIRe Project, University of Zimbabwe, Harare
[2]Department of Economics, University of Zimbabwe, Harare
Email: smartson_p@gmail.com

ABSTRACT

This piece of work uses monthly time series data on new dysentery cases at Gweru Provincial Hospital (GPH) from Janaury 2010 to December 2018, to predict dysentery cases over the period January 2019 to December 2020. As confirmed by unit root tests, the series under consideration is basically an I (1) variable. The study applied the Box-Jenkins "catch all" model. Residual analysis of this model indicates that the model is stable and thus suitable for predicting dysentery cases at GPH over the out-of-sample period. The results of the study reveal that dysentery cases will be on the rise at GPH over the out-of-sample period; characterized by seasonal repeats in December each year. The study offers a two-fold policy recommendation in order to help policy makers in the fight against dysentery in children under five years of age within the GPH catchment area.

Keywords: box, children age, hostipal, Gweru Provincial Hospital

1. INTRODUCTION

All of Zimbabwe's major urban centres are haunted by an inadequate water supply. The already bad situation is compounded by poor sanitation and hygiene that has characterized literally the whole country (Nyoni & Nyoni, 2020). This situation has resulted in the transmission of such communicable diseases as dysentery amongst others (Makwara & Tavuyanago, 2012). Dysentery is caused by various species of the Shigella bacteria which infect the large and part of small intestines and invades the cells lining of the bowel and multiplies there, killing the cells. Its symptoms include fever, abdominal pain and uncontrolled loose or water stools (Sampson & Leslie, 1979). Culture (that is, growing bacteria in the laboratory) of freshly obtained diarrhoea fluid is the only way to be certain of the diagnosis. However, this may not always be positive, especially if the patient is already on antibiotics. Severe cases of dysentery are fatal, especially among children (WHO, 2009). Dysentery is basically transmitted through faecal-oral route via contaminated food and water (Yan et al., 2017) and usually occurs in developing countries such as Zimbabwe where there is poor sanitation and unsafe water supply. This study, whose objectives are 3-fold as outlined below; seeks to model and forecast dysentery cases in children under the age of five at Gweru Provincial Hospital (GPH), which is referral hospital in the tertiary level of the health system of Zimbabwe. GPH basically serves patients reffered from district hospitals in the Midlands province.

1.1 OBJECTIVES OF THE STUDY

i. To analyze new dysentery cases in children under five years of age at GPH over the period January 2010 to December 2018.
ii. To forecast dysentery cases at GPH over the period January 2019 to December 2020.
iii. To determine whether dysentery cases are increasing or decreasing at GPH over the out-of-sample period.

1.2 RELEVANCE OF THE STUDY

In Zimbabwe, the burden of dysentery is on the rise (Nyoni & Nyoni, 2020) and yet effective methods for forecasting dysentery cases or incidences are lacking (Yan et al., 2017). In Zimbabwe, dysentery cases are usually characterised by high mortality, especially in children (Oneko et al., 2001). This warranties the need for effective dysentery forecasting models in order to improve health policy formulation (Nyoni & Nyoni, 2020). Hence, it becomes inevitable to investigate dysentery trends that have been recorded at GPH, particularly for under-fives because they are the most vulnerable group.

2. LITERATURE REVIEW

In China, in the city of Yichang, Yan et al. (2010) forecasted dysentery cases using hybrid models and found out that the SARIMA-GRNN model performs better than SARIMA models. In another Chinese study, Tang et al. (2014), focused on Jiangsu province and investigated Shigella patterns and risk factors using GIS tools and ARIMA models. Their study revealed that the ARIMA (1, 12, 0) performed very well in comparison to GIS tools. In another study done in Jiangsu province in China, Wang et al. (2016) predicted dysentery cases using SARIMA

models and concluded that the SARIMA $(1, 1, 1)(1, 1, 2)_{12}$ model was the optimal model. In a more recent Chinese study, Yan et al. (2017), investigated the impact of meteorological factors on the incidence of bacillary dysentery in Beijing, using the ARIMAX model and concluded that temperature and rainfall were positively correlated with the number of bacillary dysentery cases in Beijing. In Zimbabwe, Nyoni & Nyoni (2020) used the Artificial Neural Networks (ANN) approach to model and forecast dysentery cases in children under the age of five in Chitungwiza urban district. Their model basically reveals a downwards trajectory of new dysentery cases in children under five years of age in Chitungwiza urban district. The reviewed literature shows that the SARIMA model is generally more popular than other models when it comes to analyzing dysentery cases. This study follows the intuition of Wang et al. (2016). However, this paper will not explore other SARIMA models like what was done by Wang et al. (2016) but rather we apply the Box-Jenkins "catch all" model, which is well known for its capability to explain monthly time series data very well as compared to other models.

3. METHODOLOGY

The Box-Jenkins SARIMA model may be expressed as follows:

$$\phi_p(B)\emptyset_p(B^s)DS_t = \theta_q(B)\Theta_q(B^s)\varepsilon_t \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots[1]$$

Where B is the backshift operator, $\phi_p, \emptyset_p, \theta_q$ and Θ_q are polynomials of order p, P, q and Q respectively. ε_t is a white noise process and $DS_t = \nabla_d\Delta_s^D Y_t$ is the differenced R series. This study applies the Box-Jenkins "catch all" model.

3.1 DATA

This study is based on monthly observations of new Dysentery cases at Gweru Provincial Hospital (GPH), from January 2010 to December 2018. The out-of-sample forecast ranges over the period January 2019 to December 2020. All the data employed in this study was gathered from the Health Information department at GPH.

3.2 DIAGNOSTIC TESTS AND MODEL EVALUATION

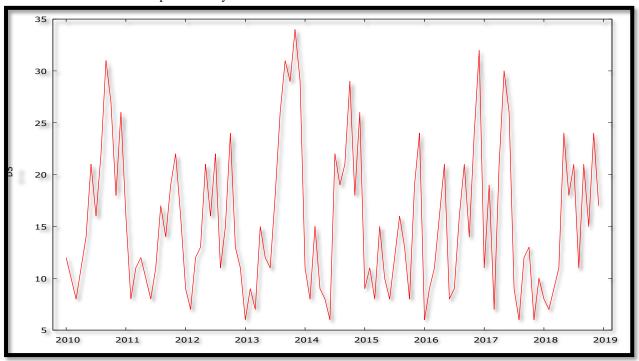Unit Root Tests: Graphical Analysis

Figure 1: Graphical Analysis

Figure 1 above shows the pattern of dysentery cases in children under 5 years of age recorded and managed at GPH. It is almost apparent that the series is stationary. However, a formal unit root test will help us verify this claim.

The ADF Test

Table 1: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
DS_t	-6.267617	0.0000	-3.492523	@1%	Stationary

| | | | -2.888669 | @5% | Stationary |
| | | | -2.581313 | @10% | Stationary |

Table 2: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
DS_t	-6.240755	0.0000	-4.046072	@1%	Stationary
			-3.452358	@5%	Stationary
			-3.151673	@10%	Stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
DS_t	-1.615588	0.0998	-2.586960	@1%	Not stationary
			-1.943882	@5%	Not stationary
			-1.614731	@10%	Stationary

Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
$D(DS_t)$	-13.52099	0.0000	-3.493129	@1%	Stationary
			-2.888932	@5%	Stationary
			-2.581453	@10%	Stationary

Table 5: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
$D(DS_t)$	-13.45451	0.0000	-4.046925	@1%	Stationary
			-3.452764	@5%	Stationary
			-3.151911	@10%	Stationary

Table 6: 1st Difference-without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
$D(DS_t)$	-13.58402	0.0000	-2.586960	@1%	Stationary
			-1.943882	@5%	Stationary
			-1.614731	@10%	Stationary

Tables 1 – 6 basically indicate that DS is an I(1) variable.

Analysis of the Residuals of the SARIMA (0, 1, 1)(0, 1, 1)$_{12}$ Model
Residual Correlogram of the SARIMA (0, 1, 1)(0, 1, 1)$_{12}$ Model
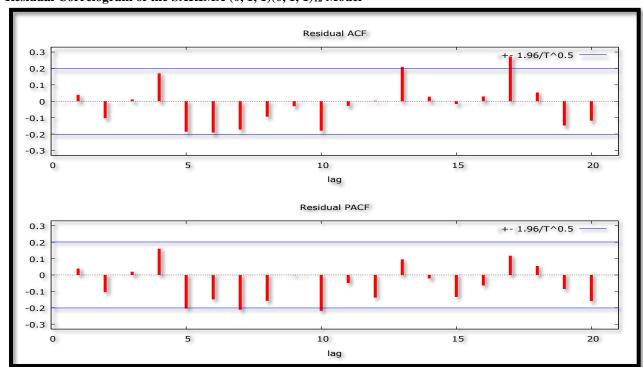

Figure 2: Residual Correlogram

Figure 2 indicates that the residuals of the Box-Jenkins "catch all" model applied in this piece of work is stable and is proof that our predictive model is suitable for forecasting monthly dysentery cases in children under the age of 5, who have been admitted and managed at GPH.

4. RESULTS OF THE STUDY

4.1 DESCRIPTIVE STATISTICS

Table 7: Summary Statistics, using the observations 2010:01 - 2019:12, for the variable DS (108 valid observations)

Mean	Median	Minimum	Maximum
15.454	14.000	6.0000	34.000
Std. Dev.	C.V.	Skewness	Ex. kurtosis
7.0966	0.45922	0.68552	-0.47124

The average number of children under 5 years of age who have suffered from dysentery over the period under study is 15; the minimum is 6 while the maximum is 34 and was recorded in November 2013.

4.2 RESULTS PRESENTATION

Table 8: Main Results of the SARIMA $(0, 1, 1)(0, 1, 1)_{12}$ Model

The SARIMA $(0, 1, 1)(0, 1, 1)_{12}$ model can be presented as follows:

$$(1 - B)(1 - B^{12})DS_t = 0.00155495 + (1 - 0.525352B)(1 - 0.744785B^{12})\varepsilon_t \ldots\ldots\ldots\ldots [2]$$

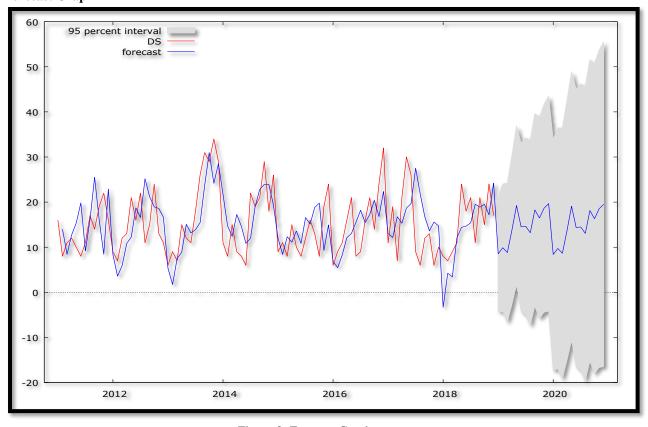| Variable	Coefficient	Standard Error	z	p-value
constant	0.00155495	0.132649	0.01172	0.9906
θ_q	-0.525352	0.0871983	-6.025	0.0000***
Θ_q	-0.744785	0.131890	-5.647	0.0000***

Forecast Graph

Figure 3: Forecast Graph

Out of Sample Forecasts

Table 9: Out-of-sample forecasts

Year: Month	Predicted DS	Standard Error	95% Confidence Interval
2019:01	8.52361	6.53120	(-4.27731, 21.3245)
2019:02	9.87793	7.22957	(-4.29177, 24.0476)
2019:03	8.81001	7.86618	(-6.60743, 24.2274)
2019:04	13.7985	8.45500	(-2.77302, 30.3700)
2019:05	19.2285	9.00539	(1.57828, 36.8788)
2019:06	14.5068	9.52404	(-4.15999, 33.1736)
2019:07	14.6393	10.0159	(-4.99145, 34.2700)
2019:08	13.1902	10.4846	(-7.35926, 33.7397)
2019:09	18.2730	10.9333	(-3.15596, 39.7019)
2019:10	16.4597	11.3643	(-5.81391, 38.7334)
2019:11	18.6400	11.7796	(-4.44748, 41.7275)
2019:12	19.6564	12.1806	(-4.21718, 43.5301)
2020:01	8.36704	13.0802	(-17.2697, 34.0037)
2020:02	9.72292	13.6467	(-17.0241, 36.4700)
2020:03	8.65655	14.1906	(-19.1566, 36.4697)
2020:04	13.6466	14.7145	(-15.1933, 42.4864)
2020:05	19.0782	15.2203	(-10.7530, 48.9094)
2020:06	14.3580	15.7098	(-16.4327, 45.1487)
2020:07	14.4920	16.1846	(-17.2291, 46.2132)
2020:08	13.0446	16.6458	(-19.5805, 45.6696)
2020:09	18.1288	17.0945	(-15.3758, 51.6335)
2020:10	16.3172	17.5318	(-18.0445, 50.6789)
2020:11	18.4990	17.9584	(-16.6989, 53.6969)
2020:12	19.5170	18.3752	(-16.4977, 55.5316)

Graphical Presentation of the Predicted Monthly Dysentery Cases in Under-fives at GPH (Out-of-Sample)
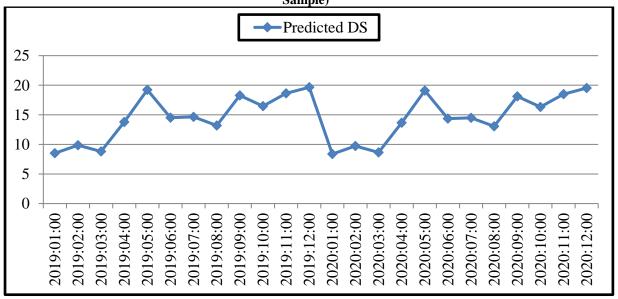

Figure 4: Graphical presentation of out-of-sample forecasts

Table 8 shows the main results of the applied model. Worthy to note is the fact that all the model parameters are statistically significant at 1% level of significance. Figures 3 and 4 as well as table 9 are out-of-sample forecasts. Predicted dysentery cases show a repeat of seasonality in December of each year. This could be attributed to high temperatures and rainfall that usually occur in December of each year in the Midlands province of Zimbabwe. These results are consistent with Yan et al. (2017) who confirmed a positive link between high temperatures & rainfall and the incidence of dysentery in China. A striking feature of figure 4, that probably raises the eyebrows of researchers and policy makers is the upward trend of the forecasts: this is a wake-up call on the

policy makers and relevant authorities on the need to be more serious in the war against dysentery in children under five years of age. However, Nyoni & Nyoni (2020) show that it is possible to win the war against dysentery, as their study indicates that, in the case of Chitungwiza urban district, dysentery cases are likely to decrease over the out-of-sample period.

4.3 RECOMMENDATIONS

i. Improvement of sanitation and hygiene should be prioritized in the Midlands province of Zimbabwe. In this regard, the government of Zimbabwe, through the relevant ministry, should make sure that people have access to safe water. This can be done by drilling more boreholes in Gweru urban area as well as Kwekwe district. This should extend to other parts of Zimbabwe to ensure a resounding win against dysentery in Zimbabwe.

ii. There is need for the government of Zimbabwe to ensure that there is enough supply of medicines at GPH for the treatment of affected children.

5. CONCLUSION

Children, especially the under-fives are not able to do proper sanitation and hygienic practices, which makes them the most vulnerable group when it comes to dysentery outbreaks. Given the poor sanitation conditions and water woes in Gweru urban, just like in any other part of Zimbabwe; it becomes instructive to carry out a study like this one. The paper employed the Box-Jenkins SARIMA model, the SARIMA $(0, 1, 1)(0, 1, 1)_{12}$ model; in order to analyze the pattern of dysentery cases at GPH over the period under study. The results of the study are a warning signal to health policy makers as they indicate a general rise in new dysentery cases at GPH in the out-of-sample period. The study is expected to go a long way in materializing the much aspired dysentery-free Zimbabwe.

REFERENCES

1. Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control, Holden Day, San Francisco.
2. Makwara, E. C., & Tavuyanago, B. (2012). Water Woes in Zimbabwe's Urban Areas in the Midst of Plenty: 2000 – Present, European Journal of Sustainable Development, 1 (2): 151 – 180.
3. Nyoni, S. P., & Nyoni, T. (2020). Using Artificial Neural Networks for Predicting New Dysentery Cases in Children Under 5 Years of Age in Chitungwiza Urban District, Zimbabwe, International Journal of Research and Development (IJRD), 5 (2): 215 – 221.
4. Oneko, M., Nyathi, M., & Doehring, E. (2001). Post-dysenteric Hemolytic Uremic Syndrome in Bulawayo, Zimbabwe, Pediatric Nephrology, 16: 1142 – 1145.
5. Sampson, C. C., & Leslie, B. (1979). Shigellosis: Analysis of 176 Isolations, Journal of National Medical Association, 71: 35 – 36.
6. Tang, F., Cheng, Y., Bao, C., Hu, C., Hu, J., Liu, W., Liang, Q., Wu, Y., Norris, J., Peng, Z., Yu, R., Shen, H., & Chen, F. (2014). Spatio-temporal Trends and Risk Factors for Shigella From 2001 to 2011 in Jiangsu Province, People's Republic of China, PLoS ONE, 9 (1): 1 – 11.
7. Wang, K., Song, W., Li, J., Lu, W., Yu, J., & Han, X. (2016). The Use of An Autoregressive Integrated Moving Average Model for Prediction of the Incidence of Dysentery in Jiangsu, China, Asia Pacific Journal of Public Health, 28 (4): 336 – 346.
8. WHO (2009). Diarrhoeal Diseases. http://www.int/vaccine/research/disease/diarrhoeal/en/index6.html ; accessed 24/03/2020.
9. Yan, L., Wang, L., Zhang, X., Li, M. Y., & He, J. (2017). Impact of Meteorological Factors on the Incidence of Bacillary Dysentery in Beijing, China: A Time Series Analysis (1970-2012), PLoS ONE, 12 (8): 1 – 13.
10. Yan, W., Xu, Y., Yang, X., & Zhou, Y. (2010). A Hybrid Model for Short-term Bacillary Dysentery Prediction in Yichang City, China, Japan Journal of Infectious Diseases, 63: 264 – 270.